

Section 12: Rare Events

Contents

12.1 Rare-event Behavior in Stochastic Models	1
12.2 Rare Events for Random Walks	1
12.3 Change of Measure	2
12.4 Large Deviations for Random Walk	4
12.5 Maximum of Random Walk	6

12.1 Rare-event Behavior in Stochastic Models

In many stochastic models, it turns out (perhaps surprisingly) that, conditional on the occurrence of a rare event, the dynamical behavior of the model that leads to the rare event becomes more and more predictable as the event becomes rarer (i.e. one can prove limit theorems describing the conditional dynamics, as the event becomes successively rarer).

12.2 Rare Events for Random Walks

We start with discussion of rare events in the setting of random walk, in part because of the simplicity of random walk, and in part because so many models are built up from random walk. Let $S_n = Z_1 + \dots + Z_n$, when the Z_i 's are iid scalar rv's. Assuming $E|Z_1| < \infty$ and choose $a > EZ_1$. Then, by the law of large numbers,

$$P(S_n > na) \rightarrow 0.$$

as $n \rightarrow \infty$. So $\{S_n > na\}$ becomes successively rarer $n \rightarrow \infty$. Note that

$$P(S_n \geq na) \geq P(Z_1 \geq a, Z_2 \geq a, \dots, Z_n \geq a) = P(Z \geq a)^n,$$

so that $P(S_n \geq na) \geq \exp(r_1 n)$, where $r_1 = -\log P(Z_1 \geq a)$. Hence, $P(S_n \geq na)$ is lower bounded by a quantity that decreases to 0 at exponential rate (in n).

Suppose now that Z_1 is *light-tailed*, so that $E \exp(\theta Z_1) < \infty$ for θ in a neighborhood of the origin. By Markov's inequality,

$$\begin{aligned} P(S_n \geq na) &= P(\theta S_n \geq \theta na) \\ &= P(\exp(\theta S_n) \geq (\theta na)) \\ &\leq \exp(-\theta na) (E \exp(\theta Z_1))^n \\ &= \exp(-\theta na + n\phi(\theta)) \end{aligned}$$

for any $\theta > 0$, for which $\phi(\theta) \triangleq \log E \exp(\theta Z_1) < \infty$. For $\theta > 0$ and small,

$$\phi(\theta) = \phi(0) + \phi'(0)\theta + \frac{\phi''(0)}{2}\theta^2 + o(\theta^2) = EZ_1\theta + O(\theta^2)$$

as $\theta \downarrow 0$, so

$$\begin{aligned} P(S_n \geq na) &= \exp(-n(\theta a - \phi(\theta))) \\ &= \exp(-n(\theta a - \theta Z_1 + O(\theta^2))) \\ &= \exp(-n\theta(a - Z_1) + nO(\theta^2)). \end{aligned}$$

Hence, by choosing $\theta > 0$ and sufficiently small, it is evident that $P(S_n \geq na) \leq \exp(-r_2 n)$, where $r_2 > 0$, so that $P(S_n \geq na)$ is upper bounded by a quantity that decreases to zero at exponential rate in n .

This suggests that there exists a constant $I(a) > 0$ for which

$$\frac{1}{n} \log P(S_n \geq na) \rightarrow -I(a) \tag{12.2.1}$$

as $n \rightarrow \infty$, when the Z_i 's are light-tailed. We will later prove this and show how to compute the rate function ($I(a) : a > 0$); the limit theorem 12.2.1 is fundamental to the study of “large deviations”, an important sub-area of probability that focuses on rare-event behavior.

Before proceeding further, note that rare events of exponential order are what one expects when the rare event is generated by “conspiratorial behavior”. In particular, when order n independent rv's must behave in an unusual manner to generate the rare event, we expect a rare event probability of exponential order. On the other hand, if the rare event can be generated by a small number of independent rv's behaving unusually, 12.2.1 will usually be violated.

Suppose, for example, that

$$P(Z_1 > x) \sim cx^{-\alpha}$$

as $x \rightarrow \infty$, for $\alpha, c > 0$. In this setting,

$$\begin{aligned} P(S_n \geq na) &\geq \sum_{i=1}^n P(Z_i \geq n(a - EZ_1), \sum_{j>i} Z_j \geq nEZ_1) \\ &= nP(Z_1 \geq n(a - EZ_1))P(S_{n-1} \geq nEZ_1) \\ &\sim \frac{n}{2} cn^{-\alpha} (a - EZ_1)^{-\alpha} \end{aligned}$$

as $n \rightarrow \infty$, so that $P(S_n \geq na)$ is of polynomial order in n , rather than exponential order in n . Similarly, whenever the Z_i 's are heavy-tailed, we can expect that 12.2.1 is violated and that the rare event $\{S_n \geq na\}$ is generated by only a small number of the Z_i 's, (rather than through conspiratorial behavior).

12.3 Change of Measure

A key tool in analyzing rare events is the use of “change-of-measure” arguments. Such arguments are intimately connected to the theory of martingales.

Given a probability P on Ω , note that if R is a non-negative rv for which $ER = 1$, then

$$Q(dw) = R(w)P(dw)$$

defines a probability on Ω and

$$E_Q W = EWR$$

for each non-negative rv W . Furthermore, if $R > 0$ a.s. (under P), then

$$EW = E_Q WR^{-1}$$

so expectations under P can be computed (via a “change-of-measure”) as expectations under Q or vice versa.

Now, consider a sequence $(R_n : n \geq 1)$ of non-negative rv’s for which $ER_n = 1$ for $n \geq 1$. Set

$$Q_n(dw) = R_n(w)P(dw).$$

In general, such a sequence of probabilities $(Q_n : n \geq 1)$ can have arbitrary structure, so that Q_n has nothing to do with Q_{n+1} , for example.

However, in the special case that $(M_n : n \geq 0)$ is a nonnegative martingale adopted to the sequence $(Z_n : n \geq 0)$, having $EM_n = 1$, the corresponding probabilities

$$\tilde{P}_n(dw) = M_n(w)P(dw)$$

are intimately connected. In this case, note that

$$\begin{aligned} & \tilde{P}_{n+1}((Z_0, \dots, Z_n) \in A) \\ &= \tilde{E}_{n+1} \mathbf{1}((Z_0, \dots, Z_n) \in A) \\ &= E \mathbf{1}((Z_0, \dots, Z_n) \in A) M_{n+1} \\ &= E \mathbf{1}((Z_0, \dots, Z_n) \in A) E[M_{n+1} | Z_0, \dots, Z_n] \\ &= E \mathbf{1}((Z_0, \dots, Z_n) \in A) M_n \\ &= \tilde{P}_n((Z_0, \dots, Z_n) \in A); \end{aligned}$$

hence the probability assignments made by the \tilde{P}_n ’s are consistent. The Kolmogorov extension theorem then guarantees existence of a probability \tilde{P} for which

$$\tilde{P}((Z_0, \dots, Z_n) \in A) = E \mathbf{1}((Z_0, \dots, Z_n) \in A) M_n$$

for each $n \geq 0$. Thus, there exists a single probability \tilde{P} that “knits together” all the \tilde{P}_n ’s.

Let us illustrate this idea in the setting of the exponential martingale

$$M_n(\theta) = \exp(\theta S_n - n\phi(\theta))$$

(that is adapted to $(Z_n : n \geq 0)$). The Kolmogorov extension theorem guarantees existence of a probability \tilde{P}_θ for which

$$\begin{aligned} & \tilde{P}_\theta((Z_i \in A_i, 1 \leq i \leq n)) \\ &= E \prod_{i=1}^n \mathbf{1}(Z_i \in A_i) \exp(\theta \sum_{i=1}^n Z_i - n\phi(\theta)) \\ &= \prod_{i=1}^n E \mathbf{1}(Z_i \in A_i) \exp(\theta Z_i - \phi(\theta)) \\ &= \prod_{i=1}^n \tilde{P}_\theta(Z_i \in A_i), \end{aligned}$$

where

$$\tilde{P}_\theta(Z_1 \in dz) = e^{\theta z - \phi(\theta)} P(Z_1 \in dz) \quad (12.3.1)$$

for $z \in \mathbf{R}$. Hence, \tilde{P}_θ is a distribution under which the Z_i 's continue to be iid, but with marginal distribution “exponentially tailed” as defined by 12.3.1. In fact,

$$\tilde{E}_\theta Z_1 = \frac{E Z_1 e^{\theta Z_1}}{E e^{\theta Z_1}} = \phi'(0)$$

and

$$\tilde{\text{var}}_\theta Z_1 = \tilde{E}_\theta Z_1^2 - (\tilde{E}_\theta Z_1)^2 = \phi''(0).$$

12.4 Large Deviations for Random Walk

We now apply the change-of-measure idea to computing $P(S_n \geq na)$ for n large. Note that

$$\begin{aligned} S_n &= E[S_n | S_n] \\ &= \sum_{i=1}^n E[Z_i | S_n] \\ &= nE[Z_1 | S_n], \end{aligned}$$

so $E[Z_1 | S_n] = S_n/n$. On $\{S_n \geq na\}$, we expect $S_n \approx na$ (for otherwise, the event will be even rarer), and hence we believe that

$$E[Z_1 | S_n \geq na] \rightarrow a$$

as $n \rightarrow \infty$.

Given this understanding of the problem, let's use the change-of-measure \tilde{P}_{θ_a} to compute $P(S_n \geq na)$, where θ_a is chosen so that

$$\tilde{E}_{\theta_a} Z_1 = a.$$

In other words we will choose $\theta_a > 0$ so that $\phi'(\theta_a) = a$. Since $M_n(\theta_a)$ is a strictly positive martingale:

$$\begin{aligned} P(S_n \geq na) &= \tilde{E}_{\theta_a} \mathbf{1}(S_n \geq na) M_n(\theta_a)^{-1} \\ &= \tilde{E}_{\theta_a} \mathbf{1}(S_n \geq na) \exp(-\theta_a S_n + n\phi(\theta_a)) \\ &= \exp(-\theta_a na + n\phi(\theta_a)) \tilde{E}_{\theta_a} \exp(-\theta_n(S_n - na)) \mathbf{1}(S_n \geq na). \end{aligned}$$

But

$$\begin{aligned} &\tilde{E}_{\theta_a} \exp(-\theta_n(S_n - na)) \mathbf{1}(S_n \geq na) \\ &= \tilde{E}_{\theta_a} \int_{S_n - na}^{\infty} \theta_a e^{-\theta_a y} dy \mathbf{1}(S_n \geq na) \\ &= \tilde{E}_{\theta_a} \int_0^{\infty} \theta_a e^{-\theta_a y} \mathbf{1}(0 \leq S_n - na \leq y) dy \\ &= \int_0^{\infty} \theta_a e^{-\theta_a y} \tilde{P}_{\theta_a}(0 \leq S_n - na \leq y) dy \\ &= \int_0^{\infty} \theta_a e^{-\theta_a y} \tilde{P}_{\theta_a} \left(0 \leq \frac{S_n - na}{\sqrt{n}} \leq \frac{y}{\sqrt{n}} \right) dy \end{aligned} \quad (12.4.1)$$

Since the Z_i 's are iid with mean a under \tilde{P}_{θ_a} ,

$$\frac{S_n - na}{\sqrt{n}} \Rightarrow \sqrt{\phi''(\theta_a)}N(0, 1)$$

as $n \rightarrow \infty$ under \tilde{P}_{θ_a} . This suggests that

$$\begin{aligned} & \tilde{P}_{\theta_a} \left(0 \leq \frac{S_n - na}{\sqrt{n}} \leq \frac{y}{\sqrt{n}} \right) \\ & \approx P \left(0 \leq \sqrt{\phi''(\theta_n)}N(0, 1) \leq \frac{y}{\sqrt{n}} \right) \\ & \approx \frac{1}{\sqrt{2\pi\phi''(\theta_a)}} \cdot \frac{y}{\sqrt{n}} \end{aligned} \tag{12.4.2}$$

as $n \rightarrow \infty$. If we plug this into 12.4.1, we arrive at

$$\begin{aligned} & \tilde{E}_{\theta_a} \exp(-\theta_a(S_n - na)\mathbf{1}(S_n \geq na)) \\ & \approx \frac{1}{\sqrt{2\pi\phi''(\theta_a)}} \int_0^\infty \theta_a y e^{-\theta_a y} dy \cdot \frac{1}{\sqrt{n}} \\ & = \frac{1}{\theta_a} \frac{1}{\sqrt{2\pi n \phi''(\theta_a)}}, \end{aligned} \tag{12.4.3}$$

and hence

$$P(S_n \geq na) \sim e^{-nI(a)} \frac{1}{\theta_a} \frac{1}{2\pi n \phi''(\theta_a)}$$

as $n \rightarrow \infty$, where $I(a) = \theta_a a - \phi(\theta_a)$. Making this asymptotic rigorous requires two additional steps:

- Observe that the central limit theorem only asserts that for each $a < b$,

$$\tilde{P}_{\theta_a} \left(a \leq \frac{S_n - na}{\sqrt{n}} \leq b \right) \rightarrow P \left(a \leq \sqrt{\phi''(\theta_a)}N(0, 1) \leq b \right)$$

as $n \rightarrow \infty$, and hence it only allows one to conclude that

$$\tilde{P}_{\theta_a} \left(0 \leq \frac{S_n - na}{\sqrt{n}} \leq \frac{y}{\sqrt{n}} \right) \rightarrow 0$$

as $n \rightarrow \infty$. To obtain 12.4.2 requires the application of the *local central limit theorem* and demands that the Z_i 's have a density.

- We need to show that the Dominated Convergence Theorem applies at step 1.

With these two additional steps, we arrive at the following theorem.

Theorem 12.4.1 Suppose that the Z_i 's are iid random variables for which their common distribution has a density. If there exists $\theta_a > 0$ at which $\phi'(\theta_a) = a > EZ_1$ and ϕ is twice differentiable at θ_a , then

$$P(S_n \geq na) \sim e^{-nI(a)} \frac{1}{\theta_a} \frac{1}{2\pi n \phi''(\theta_a)} \tag{12.4.4}$$

as $n \rightarrow \infty$.

Remark 12.4.1 The dominant term in the above approximation is $\exp(-nI(a))$. In fact, 12.4.4 implies that

$$\frac{1}{n} \log P(S_n \geq na) \rightarrow -I(a)$$

as $n \rightarrow \infty$, so 12.4.4 shows that the rate function $I(a)$ is given by

$$I(a) = a\theta_a - \phi(\theta_a).$$

An interesting observation is that

$$I(a) = \sup_{\theta} [\theta a - \phi(\theta)] = \phi^*(a),$$

so that $I(\cdot)$ is the *convex conjugate* (or “convex dual”) of the convex function ϕ . Convexity plays a major role in the theory of large deviations.

Remark 12.4.2 With some additional effort, one can prove that

$$P(Z_1 \in A, \dots, Z_n \in A_n | S_n \geq an) \rightarrow \tilde{P}_{\theta_a}(Z_1 \in A_1, \dots, Z_n \in A_n)$$

as $n \rightarrow \infty$, so that $\tilde{P}_{\theta_a}(\cdot)$ can be interpreted, in some sense, as the conditional distribution of the Z_i 's under the conditioning $\{S_n \geq na\}$. This, in turn, can be viewed as a mathematical elaboration of the fact that it is conspiratorial behavior on the part of the Z_i 's that leads to $\{S_n \geq na\}$ when the Z_i 's are light-tailed.

12.5 Maximum of Random Walk

As we know,

$$M_\infty = \max_{k \geq 0} S_k$$

arises as the limiting distribution of the waiting time sequence for the FIFO G/G/1 queue (when the queue is stable, so that $EZ_1 < 0$). We will now show how to develop an approximation for the tail probabilities

$$P(M_\infty > x)$$

when x is big and the Z_i 's are light-tailed. Observe that

$$P(M_\infty > x) = P\left(\max_{k \geq 0} S_k > x\right).$$

On the other hand,

$$\begin{aligned} P(S_k > x) &= P(S_{\lfloor x/t \rfloor} > x) \quad (t = k/x) \\ &\approx \exp\left(-x_t I\left(\frac{1}{t}\right)\right). \end{aligned}$$

Let

$$t^* = \arg \min_{t > 0} t I\left(\frac{1}{t}\right).$$

Using the inequality

$$P(S_{\lfloor x/t^* \rfloor} > x) \leq P\left(\max_{k \geq 0} S_k > x\right) \leq \sum_{k=1}^{\infty} P(S_k > x),$$

it is not difficult to establish, in significant generality, that

$$\frac{1}{x} \log P \left(\max_{k \geq 0} S_k > x \right) \rightarrow -t^* I \left(\frac{1}{t^*} \right)$$

as $n \rightarrow \infty$. To simplify this answer, note that

$$\begin{aligned} & \frac{d}{dt} \left(t I \left(\frac{1}{t} \right) \right) \\ &= \frac{d}{dt} \left(t \theta_{\frac{1}{t}} \left(\frac{1}{t} \right) - t \phi \left(\theta_{\frac{1}{t}} \right) \right) \\ &= \frac{d}{dt} \left(\theta_{\frac{1}{t}} - t \phi \left(\theta_{\frac{1}{t}} \right) \right) \\ &= \theta'_{\frac{1}{t}} \left(-\frac{1}{t^2} \right) - t \phi' \left(\theta_{\frac{1}{t}} \right) \theta'_{\frac{1}{t}} \left(-\frac{1}{t^2} \right) - \phi \left(\theta_{\frac{1}{t}} \right) \\ &= \theta'_{\frac{1}{t}} \left(-\frac{1}{t^2} \right) - \theta'_{\frac{1}{t}} \left(-\frac{1}{t^2} \right) - \phi \left(\theta_{\frac{1}{t}} \right) \\ &= \phi \left(\theta_{\frac{1}{t}} \right), \end{aligned}$$

so the optimizer t^* is such that $\phi \left(\theta_{\frac{1}{t^*}} \right) = 0$. Consequently,

$$t^* I \left(\frac{1}{t^*} \right) = \theta_{\frac{1}{t^*}}$$

and hence

$$\frac{1}{x} \log P \left(\max_{k \geq 0} S_k > x \right) \rightarrow -\theta^*$$

as $x \rightarrow \infty$, where $\theta^* > 0$ is such that $\phi(\theta^*) = 0$.

We can now improve upon this result by using our “change-of-measure” ideas. Note that

$$P \left(\max_{k \geq 0} S_k > x \right) = P(T_x < \infty)$$

, where $T_x = \inf\{n \geq 0 : S_n > x\}$. Then,

$$\begin{aligned} P(T_x = n) &= \tilde{E}_{\theta^*} \mathbf{1}(T_x = n) M_n^{-1}(\theta^*) \\ &= \tilde{E}_{\theta^*} \mathbf{1}(T_x = n) \exp(-\theta^* S_n) \\ &= \tilde{E}_{\theta^*} \mathbf{1}(T_x = n) \exp(-\theta^* S_{T_x}) \end{aligned}$$

So

$$P(T_x < \infty) = \tilde{E}_{\theta^*} \mathbf{1}(T_x < \infty) \exp(-\theta^* S_{T_x})$$

Because $\phi(\cdot)$ is convex and $\phi'(0) = EZ_1 < 0$, evidently $\phi'(\theta^*) > 0$. Hence, $\{T_x < \infty\}$ is a probability one event under \tilde{P}_{θ^*} , so

$$P(T_x < \infty) = \tilde{E}_{\theta^*} \exp(-\theta^* S_{T_x}) = \exp(-\theta^* x) \tilde{E}_{\theta^*} \exp(-\theta^*(S_{T_x} - x)).$$

Hence,

$$P(T_x < \infty) \leq \exp(-\theta^* x)$$

for $x \geq 0$; this is the celebrated Cramer-Lundberg inequality.

To proceed further, let

$$\beta_n = \inf\{j > \beta_{n-1} : S_j > S_{\beta_{n-1}}\}$$

be the n 'th time at which the random walk attains a new maximum under \tilde{P}_{θ^*} . Then $(S_{\beta_n} : n \geq 0)$ is a strictly increasing sequence, and

$$S_{T_x} = S_{\beta_{N(x)+1}},$$

where $N(x) + 1 = \inf\{j \geq 0 : S_{\beta_j} > x\}$. Because $(S_{\beta_j} - S_{\beta_{j-1}} : j \geq 1)$ is an iid sequence of positive rv's. We can apply a renewal argument to analyze $E_{\theta^*} \exp(-\theta^*(S_{T_x} - x))$:

$$\tilde{E}_{\theta^*} \exp(-\theta^*(S_{T_x} - x)) = \tilde{E}_{\theta^*} \exp(-\theta^*(S_{\beta_1} - x)) \mathbf{1}(S_{\beta_1} > x) + \int_{[0,x]} \tilde{E}_{\theta^*} \exp(-\theta^*(S_{T_{x-y}} - (x-y))) P_{\theta^*}(S_{\beta_1} \in dy)$$

If the Z_j 's have a density, S_{β_1} is spread-out so Smith's version of the renewal theorem applies:

$$\tilde{E}_{\theta^*} \exp(-\theta^*(S_{T_x} - x)) \rightarrow c \triangleq \frac{1}{\theta^*} E_{\theta^*}(1 - \exp(-\theta^* S_{\beta_1})) / E_{\theta^*} S_{\beta_1}$$

as $n \rightarrow \infty$. Hence, we conclude that

$$P\left(\max_{k \geq 0} S_k > x\right) \sim c \exp(-\theta^* x)$$

as $x \rightarrow \infty$, this is known as the Cramer-Lundberg asymptotic for the tail of the maximum of random walk.

Remark 12.5.1 It is easy to show here that

$$P\left(Z_1 \in A_1, \dots, Z_m \in A_m \mid \max_{k \geq 0} S_k > x\right) \rightarrow P_{\theta^*}(Z_1 \in A_1, \dots, Z_m \in A_m)$$

as $n \rightarrow \infty$. Again, this establishes that in the light-tailed setting, $\{\max_{k \geq 0} S_k > x\}$ occurs as a consequence of conspiratorial behavior.